



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

A global-local artificial neural network with application to wave overtopping prediction

Citation for published version:

Wedge, DC, Ingram, D, Mingham, CG, McLean, D & Bandar, ZA 2005, A global-local artificial neural network with application to wave overtopping prediction. in *Artificial Neural Networks: Formal Models and Their Applications – ICANN 2005: 15th International Conference, Warsaw, Poland, September 11-15, 2005. Proceedings, Part II*. vol. 3697, Lecture Notes in Computer Science, vol. 3697, Springer-Verlag GmbH, pp. 109-114. https://doi.org/10.1007/11550907_18

Digital Object Identifier (DOI):

[10.1007/11550907_18](https://doi.org/10.1007/11550907_18)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Early version, also known as pre-print

Published In:

Artificial Neural Networks: Formal Models and Their Applications – ICANN 2005

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



A Global-Local Artificial Neural Network with Application to Wave Overtopping Prediction

David Wedge, David Ingram, David McLean, Clive Mingham and Zuhair Bandar

Department of Computing and Mathematics, Manchester Metropolitan University, John Dalton building, Chester Street, Manchester, M1 5GD

Abstract. We present a hybrid Radial Basis Function (RBF) - sigmoid neural network with a three-step training algorithm that utilises both global search and gradient descent training. We test the effectiveness of our method using four synthetic datasets and demonstrate its use in wave overtopping prediction. It is shown that the hybrid architecture is often superior to architectures containing neurons of a single type in several ways: lower errors are often achievable using fewer hidden neurons and with less need for regularisation. Our Global-Local Artificial Neural Network (GL-ANN) is also seen to compare favourably with Perceptron Radial Basis Net (PRBFN).

1 Introduction

Multi-Layer Perceptron (MLP) and RBF networks have complementary properties. While both are theoretically capable of approximating a function to arbitrary accuracy using a single hidden layer, their operation is quite different [1]. MLP networks have a fixed architecture and are usually trained using a variant of gradient descent. In most cases, stochastic rather than batch training gives better results. MLP networks invariably incorporate neurons with sigmoid activation functions. Their response therefore varies across the whole input space and weight training is affected by all training points. RBF networks, on the other hand, are most commonly created using a constructive algorithm. Gradient descent training is usually replaced by deterministic, global methods such as Forward Selection of Centres with Orthogonal Least Squares (FS-OLS). Unlike sigmoid neurons, RBF neurons respond strongly only to inputs within a local region.

We present a hybrid network that combines the global approximation capabilities of MLP networks with the local approximation capabilities of RBF networks. It is tested using 4 synthetic datasets and comparisons are made with alternative architectures and training methods, including PRBFN [2]. Our network is then applied to the real-world problem of wave overtopping prediction.

2 Global-Local Artificial Neural Networks

A hybrid ANN containing both sigmoidal and radial neurons has the advantages of both RBF and MLP ANNs. The approach in PRBFN is to cluster the data and to

choose a neuron, either sigmoidal or radial-based, that approximates the local function within each cluster [2]. We approximate on a global level first using a MLP and then add RBF neurons using FS-OLS, in order to add local detail to the approximating function. For this reason we call our network a Global-Local Artificial Neural Network (GL-ANN). The advantages of our approach are-

- There is no need to cluster the data prior to training. This gives more flexibility to the FS-OLS process and avoids possible problems when clustering reflects the distribution of the available data rather than the underlying functionality.
- All phases of training take into account all of the training data.
- Unlike pure RBF networks and PRBFNs, our networks do not require regularisation. The MLP created in the first phase of training has a moderating effect on the selection and training of RBF neurons added subsequently.

3 Training method

In order to achieve rapid training, the Levenberg-Marquardt (L-M) method is used to train the MLP networks. Fixed numbers of sigmoidal neurons are used in a single hidden layer and the output neuron has a linear activation function. Initial weights are set to small random values at the start of training and inputs and outputs undergo a linear transformation to give them a range of $[-0.8, 0.8]$.

RBF neurons are then added using FS to choose RBF centres from the training data. After each addition the output weights from both sigmoidal and RBF neurons are adjusted using OLS minimisation. Symmetrical radial functions with fixed widths are employed. Finally all weights, including hidden layer weights and RBF steepnesses, are optimised using L-M training. Using this algorithm a series of networks with different architectures may be created. Their performance is then assessed using unseen test data. In each case the data is sampled several times to determine the training - test split and averages are taken over all runs. 10 runs are used for the benchmark tests and 30 for the wave overtopping data.

4 Benchmark tests

4.1 Method and data-sets

Four benchmark tests are employed. They are all function approximation tasks using synthetic data. The tests are taken from [2], where comparisons are made with a number of other approaches. For this reason the treatment varies between the different tests. While this creates some inconsistency it allows the consideration of a variety of datasets and permits comparison with a number of alternative methods.

The first function is

$$f(x)=\sin(12x) \quad (1)$$

with x randomly selected from $[0,1]$ and $f(x)$ corrupted by Gaussian noise of standard deviation 0.1. 50 samples are used for training and 50 for testing [3]. Given the noisy data, the minimum Mean Square Error (MSE) achievable is 0.01.

The second function is the 2D sine wave

$$f(x)=0.8\sin(x_1/4)\sin(x_2/2) \quad (2)$$

with $x_1 \in [0,10]$ and $x_2 \in [-5,5]$. The training data is randomly selected and again corrupted with Gaussian noise of standard deviation 0.1, but clean data is used for testing purposes, arranged in a 20 by 20 grid to cover the entire input space. 200 samples are used for training and 400 samples for testing.

The third function is a simulated alternating current used by Friedman in the evaluation of multivariate adaptive regression splines (MARS) [4]. It is given by

$$Z(R,\omega,L,C)=\sqrt{(R^2+(\omega L-1/\omega C)^2)} \quad (3)$$

where Z is the impedance, R the resistance, ω the angular frequency, L the inductance and C the capacitance of the circuit. The input ranges are $R \in [0,100]$, $\omega \in [40\pi, 560\pi]$, $L \in [0,1]$ and $C \in [1 \times 10^{-6}, 11 \times 10^{-6}]$. 200 random samples with Gaussian noise of standard deviation 175 added to Z are used for training. 5000 random clean samples are used for testing.

The fourth function is the Hermite polynomial,

$$f(x)=(1+(x+2x^2))\exp(-x^2) \quad (4)$$

with x randomly selected from $[-4,4]$. 100 random samples corrupted by Gaussian noise of standard deviation 0.1 are used for training purposes. 100 clean samples are used for testing. This function was first used by Mackay [5].

4.2 Results and Discussion

Mean MSE results averaged over 10 runs for MLP, FS-OLS and GL-ANN networks are given in Table 1. In the case of the third data-set we follow Friedman [4] in dividing the MSE by the variance of the test data.

Table 1. Mean Square Errors for 4 Benchmark Datasets

	1D sine	2D sine	Friedman	MacKay
MLP (L-M)	1.75e-2	1.28e-3	1.02e-1	2.14e-3
FS-OLS	1.19e-2	0.79e-3	1.55e-1	1.41e-3
GL-ANN	1.20e-2	1.11e-3	0.98e-1	1.30e-3

In the case of the more complex datasets (Friedman and Mackay) the GL-ANN is better than both pure networks. With the 1D sine data GL-ANN and FS-OLS achieve comparable results. The sine 2D dataset gives best results using a pure RBF network (FS-OLS).

Table 2 gives the number of neurons used in the most successful networks. In each case S, R and T refer to the number of sigmoid, RBF and total neurons in the hidden layer, respectively. For the first two datasets, the GL-ANN imitates the RBF networks, using just 1 sigmoid neuron. The GL-ANN uses just 3 hidden neurons to reproduce Mackay's function and 6 for Friedman's. These results show that the GL-ANN is parsimonious in its use of hidden neurons.

Table 2. Number of hidden layer neurons for 4 benchmark datasets

	1D sine			2D sine			Friedman			MacKay		
	S	R	T	S	R	T	S	R	T	S	R	T
MLP(L-M)	12	0	12	5	0	5	5	0	5	14	0	14
FS-OLS	0	6	6	0	14	14	0	43	43	0	7	7
GL-ANN	1	6	7	1	16	17	3	3	6	1	2	3

It is observed that the introduction of regularisation into OLS training of GL-ANNs does not improve the errors of the most effective networks. The initial MLP appears to have a moderating effect on the weights of RBF neurons added subsequently. Similarly, the final L-M optimisation step is found to have a beneficial effect on the GL-ANNs but leads to over-fitting with pure RBF networks.

4.3 Comparison with PRBFN

Table 3 shows the test MSEs achieved by PRBFN. For all but the 1D sine data these errors are lower than those produced by GL-ANN. Further, the results quoted for the first dataset are below the minimum error achievable, given the inbuilt noise in the data. We believe that this result cannot therefore be taken at face value.

Table 3. MSEs for PRBFN on 4 benchmark tests

	1D sine	2D sine	Friedman	MacKay
PRBFN	0.66e-2	1.28e-3	1.5e-1	1.5e-3
Minimum error	1.00e-2			

5 Wave overtopping

5.1. Introduction

Much research has been conducted into predicting overtopping at sea-walls during storm events. One approach is to use scale models in laboratories, [6], but this is expensive and time-consuming. An alternative is to numerically model a particular seawall configuration and sea-state, e.g. [7]. However, accurate simulation requires a detailed knowledge of both the geometry of the seawall and sea conditions. Results may therefore be applied only to individual scenarios.

As part of the European CLASH project [8] a large overtopping database has been compiled. We have used this database as a resource for testing our GL-ANN.

5.2. Data Characteristics and Pre-processing

10 parameters are selected for training primarily on the basis of information content [9]. The logarithm of the output parameter q_0 (mean dimensionless overtopping rate) is used in order to give a near-Normal partial distribution. The inverse of the input T_0 (dimensionless wave period) is used for the same reason. From previous research [6] it is known that the parameters R_0 (dimensionless freeboard) and T_0 have the most influence on q_0 , and that the three quantities are related approximately by the equation

$$q_0 = AT_0 \exp(-BR_0/T_0) \quad (5)$$

where A and B are empirically determined parameters that depend on the structure of the sea-wall. A and B vary slowly with changes in the structure. Further, B is of the order of 3000 times the size of A . Treating A and B as constants therefore gives the approximately linear relationship

$$\ln(q_0) \cong -BR_0/T_0. \quad (6)$$

A hybrid network could be well-suited to this data, since the MLP network may represent the approximate relationship in equation (6) well, leaving the RBF neurons to identify local variations in the function.

6 Wave Overtopping Results and Discussion

Due to the large size of the GL-ANNs, the final optimisation step is impractically slow and has been omitted. The GL-ANN and RBF networks give lower MSEs than the MLP network, but require more hidden neurons (Table 4). The GL-ANNs are superior to the pure RBF networks in terms of both MSE and hidden layer size. As

neurons are added, the errors of both networks decrease before reaching a minimum. However, the GL-ANN networks require considerably fewer neurons to achieve a given test error (Fig 1). They also give errors comparable to those from numerical simulation, even though the latter is specific to a particular structure and sea-state.

Table 3. Mean Square Errors and Hidden Layer Sizes using overtopping data

	Average test MSE	Hidden layer size
MLP (L-M)	1.16e-3	14
FS-OLS	1.04e-3	195
GL-ANN	0.99e-3	189

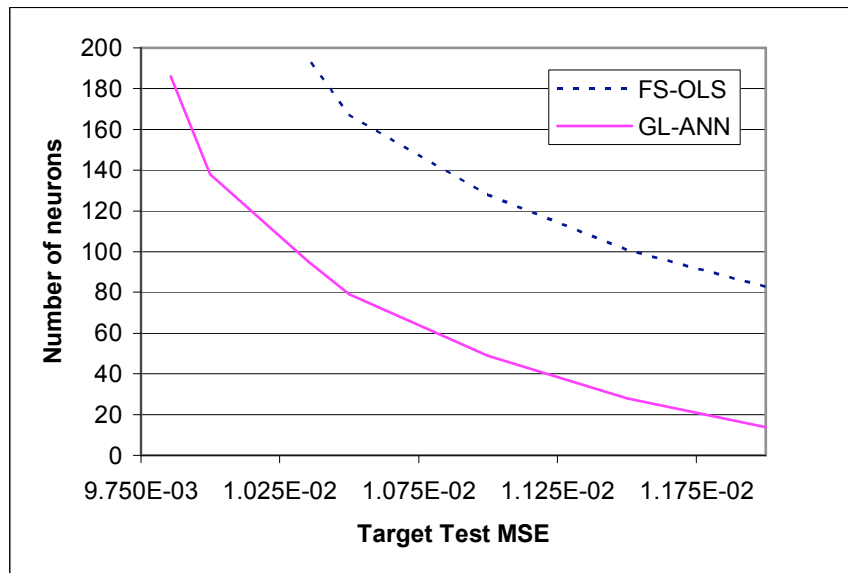


Fig. 1. Number of neurons required to achieve a given test error with FS-OLS and GL-ANN

References

1. Haykin S. *Neural Networks: A Comprehensive Foundation*. 2nd edn. Prentice Hall (1999)
2. Cohen S., Intrator N. A Hybrid Projection-based and Radial Basis Function Architecture. *Pattern Analysis and Applications*, 5: 113-120 (2002)
3. Orr M. Introduction to Radial Basis Function Networks. Available from <http://www.anc.ed.ac.uk/~mjo/rbf.html> (1996)
4. Friedman J., Multivariate adaptive regression splines. *Annals of Statistics*, 19:1-141, 1991
5. Mackay D., Bayesian Interpolation. *Neural Computation*, 4(3): 415-447 (1992)
6. Besley P. Overtopping of Seawalls: Design and Assessment Manual, Environment Agency R&D Technical Report W178 (1999).
7. Shiach J., Mingham C., Ingram D., Bruce T. The applicability of the shallow water equations for modelling violent wave overtopping. *Coastal Engineering*, 51:1-15 (2004)
8. <http://www.clash-eu.org>
9. D. Wedge, D. Ingram, C. Mingham, D. McLean, Z. Bandar. Neural Network Architectures and Overtopping Predictions. Submitted to *Maritime Engineering* (2004)